

Corner Detection, SIFT, SURF

Ing. Marek Hrúz Ph.D.
Katedra kybernetiky

Plzeň
4. října 2016

1 Harris corner detection

Roh v obraze může být definován jako průsečík dvou různých hran. Rovněž můžeme říct, že je to bod, ve kterém existují dva dominantní směry (gradienty) v lokálním okolí. Pro přesnost uvedeme, že existuje pojem významný bod. Každý roh je významný bod, ale naopak to neplatí. Významný bod může být i izolovaný bod (na pozadí jiné barvy), konec čáry, střed kružnice, atd. Algoritmus, který detekuje rohy musí být velice robustní. Tzn. že musí najít stejné rohy v podobných obrázcích, které jsou transformovány, nebo zašuměny.

1.1 Moravec corner detector

Jeden z prvních algoritmů pro detekci rohů, byl navržen Moravcem. Tento algoritmus testuje pro každý pixel podobnost okolí centrovaného na daném pixelu s okolími, které jsou poblíž. Tyto okolí se hodně překrývají. Podobnost se měří pomocí sumy absolutních rozdílů. Když je suma malá, tak podobnost je velká. Rohy v obraze jsou definovány jako pixely s malou podobností se svým okolím. Konkrétně jsou to lokální maxima sum absolutních rozdílů.

Funkce pro výpočet vypadá následovně:

$$E(u, v) = \sum_{x,y} w(x, y) [I(x + u, y + v) - I(x, y)]^2. \quad (1)$$

Pro každé x, y v obraze I se provede výpočet hodnoty E pro směry $(u, v) = \{(1, 0), (1, 1), (0, 1), (-1, 1)\}$, w je pravoúhlé okénko, které definuje lokální okolí. Ze všech směrů se vybere ten, který dává minimální odezvu (největší podoba). Tato hodnota reprezentuje sílu rohu v daném pixelu. Po výpočtu provedeném nad všema pixely se pomocí potlačení nemaxim najdou lokální maxima.

1.2 Harris corner detection

Harris se snaží reagovat na slabá místa v Moravcově článku. V první řadě nahrazuje pravoúhlé okénko Gaussovským okénkem:

$$w(x, y) = \exp\left(-\frac{(x^2 + y^2)}{2\sigma^2}\right). \quad (2)$$

Tím je potlačen šum (i na hranicích okének). Další řešený problém, je diskretizace směrů, které analyzujeme. V Harrisově článku se proto uvažují jenom malé posuny, které jsou aproximovány Taylorovým rozvojem.

$$I(x + u, y + v) \approx I(x, y) + I_u(x, y)u + I_v(x, y)v \quad (3)$$

a po dosazení do rovnice 1 získáme

$$E(u, v) \approx \sum_{x,y} w(x, y) [I_u(x, y)u + I_v(x, y)v]^2. \quad (4)$$

1 HARRIS CORNER DETECTION

Dalšími úpravami dostaneme:

$$E(u, v) \approx \sum_{x,y} w(x, y) [u^2 I_u^2 + 2uv I_u I_v + v^2 I_v^2], \quad (5)$$

což můžeme vyjádřit v maticové formě jako:

$$E(u, v) \approx \sum_{x,y} w(x, y) \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} I_u^2 & I_u I_v \\ I_u I_v & I_v^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \quad (6)$$

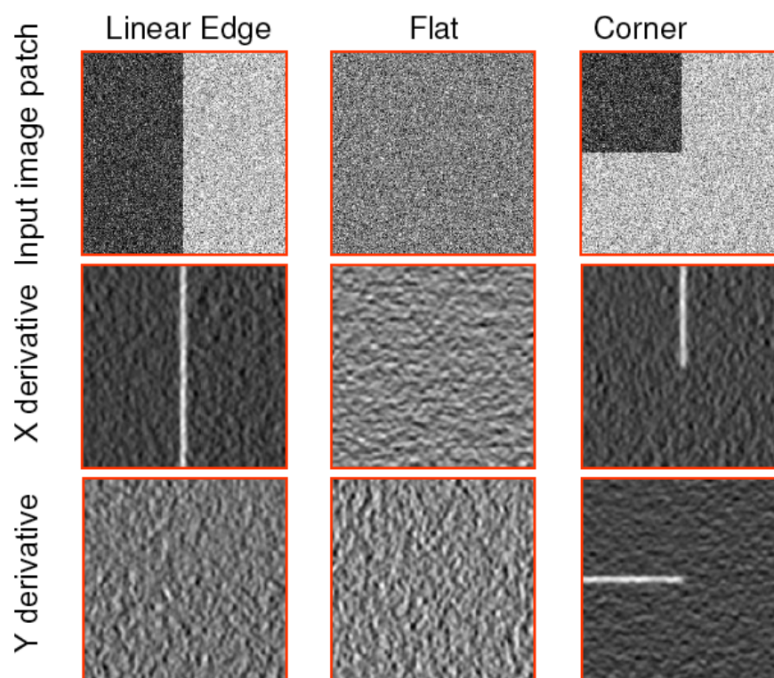
Dále definujeme

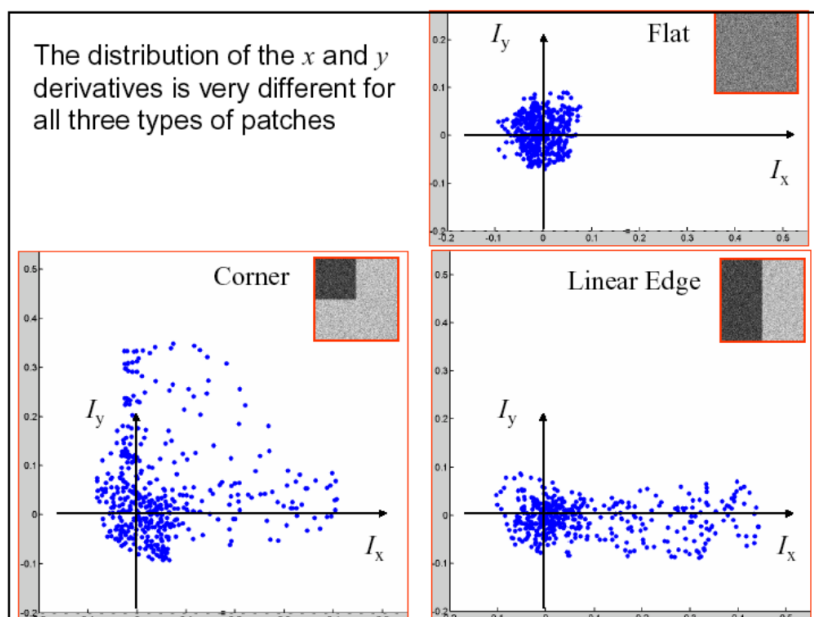
$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_u^2(x, y) & I_u I_v(x, y) \\ I_u I_v(x, y) & I_v^2(x, y) \end{bmatrix} \quad (7)$$

a můžeme psát

$$E(u, v) \approx \begin{bmatrix} u & v \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix}. \quad (8)$$

Matice M se nazývá Harrisova matice. Připomíná Hessovu matici, ale kvadratické členy jsou pouze první derivaci na druhou, kdežto v Hessově matici se jedná o druhou derivaci. Jednolivé derivace můžeme odhadnout pomocí gradientních operátorů. Okénko $w(x, y)$ vymezuje okolí, ve kterém derivace odhadujeme. Pro každý pixel máme vyčíslené M . Tato matice popisuje rozdíl mezi různě posunutými okolími analyzovaného pixelu. Analýzou této matice zjistíme o jaký typ okolí pixelu se jedná. Rozlišujeme mezi plochým regionem, hranou a rohem. Na klasifikaci použijeme vlastní čísla matice M . Abychom tento princip pochopili, budeme na problém nahlížet z jiného pohledu. Na následujícím obrázku demonstrujeme různé typy regionů.





Elipsy na Obrázku 1 reprezentují konturu Gaussovského rozložení s kovarianční maticí M . Vlastní vektory matice M odpovídají směru hlavní a vedlejší poloosy. Vlastní čísla popisují velikost těchto poloos. Podle tvaru elipsy a tedy podle velikosti vlastních čísel můžeme rozhodnout, o který případ se jedná. Jak je vidět, pokud se jedná o hranu, tak je jeden směr výraznější než druhý. To bude odpovídat situaci, kdy jedno vlastní číslo bude násobně větší než druhé. Pokud budou obě vlastní čísla porovnatelně velké, tak musíme rozhodnout o jejich absolutní velikosti. Jestli jsou příliš malé, tak se jedná o plochý región, jinak se jedná o roh. Výpočet vlastních čísel je náročná operace z hlediska výpočetních kapacit. Existuje ale možnost jak v tomto případě obejít výpočet vlastních čísel. Chceme aby vlastní čísla byla podobně velká a zároveň, aby byly obě dostatečně velké. Chceme, aby:

$$2\lambda_1\lambda_2 = \lambda_1^2 + \lambda_2^2, \quad (9)$$

což znamená, že obě vlastní čísla jsou stejná. Dále chceme, aby byli obě strany co největší.

Víme, že platí:

$$\begin{aligned} \lambda_1\lambda_2 &= \det(M) \\ \lambda_1 + \lambda_2 &= \text{trace}(M), \end{aligned} \quad (10)$$

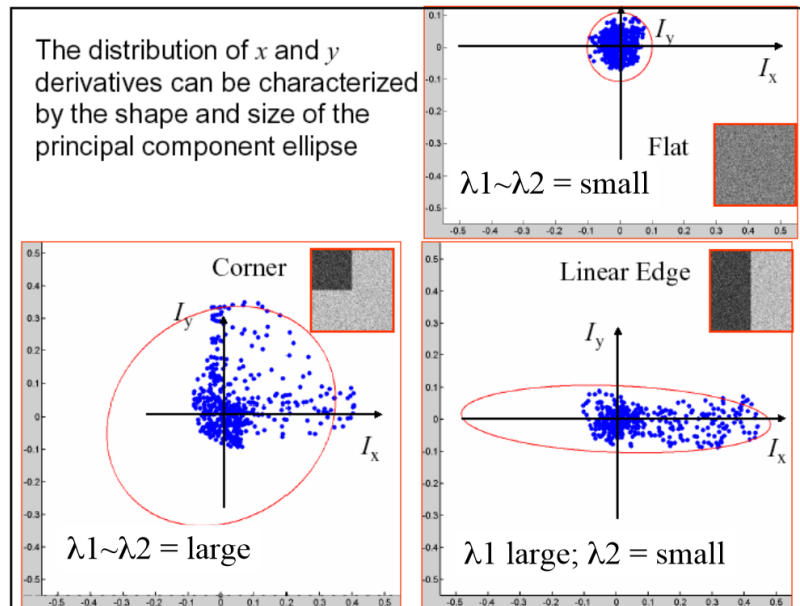
kde λ_n je vlastní číslo, $\det(M)$ je determinant matice M , $\text{trace}(M)$ je stopa matice M . Harris navrhuje vytvořit jedno měření, které by zohlední obě podmínky.

$$R = \det(M) - k(\text{trace}(M))^2, \quad (11)$$

kde k je malá konstanta (0.04...0.1). Když analyzujeme R , tak zjistíme:

- velké $R > 10000$, tak se jedná roh
- záporné a velké $R < -10000$, tak se jedná o hranu

1 HARRIS CORNER DETECTION



Obrázek 1: Body proložené elipsami.

- malé $R \in (-10000; 10000)$, tak se jedná o plochý región

2 SIFT

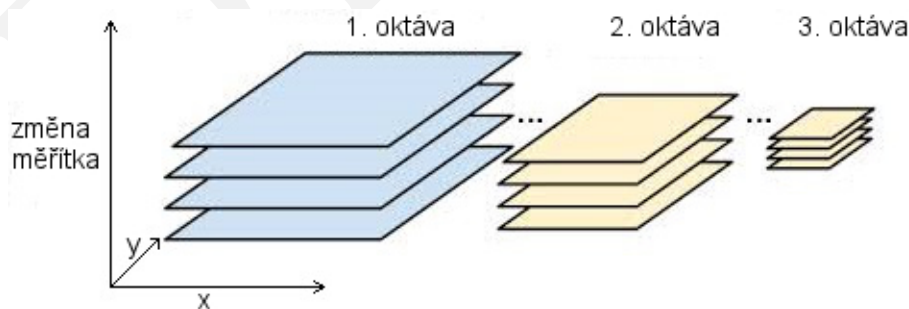
Metoda SIFT (z ang. *Scale-Invariant Feature Transform*) byla poprvé použita za účelem detekce objektů v obrazové scéně. Z trénovací množiny obrazů byly vypočteny příznakové vektory objektů, které byly následně hledány v obrazech testovacích. Pokud si vektory získané při trénování a testování dostatečně odpovídaly, byl objekt detekován a zároveň rozpoznán.

Tento princip lze ovšem stejně dobře využít i při klasifikaci. Z trénovacích množin (jedna pro každou třídu) jsou algoritmem získány příznakové vektory a ty jsou následně porovnávány s vektory spočtenými pro testovací množinu. V dalším kroku je pomocí zvolené klasifikační metody rozhodnuto o rozdělení prvků testovací množiny do jednotlivých tříd.

SIFT se skládá ze čtyř hlavních kroků: (1) detekce extrémů uvnitř scale-space; (2) zpřesnění polohy významných bodů; (3) přiřazení orientace významným bodům; (4) sestavení deskriptoru významných bodů.

DETEKCE EXTRÉMŮ UVNITŘ SCALE-SPACE

Detaily v obraze jsou tvořeny strukturami s malým měřítkem, pokud měřítko postupně zvětšujeme, zachováváme v obraze pouze ty struktury, které mají měřítko vyšší. Diskrétní scale-space je 3D reprezentace obrazu o několika vrstvách, kde první vrstva odpovídá původnímu obrazu a každá další je tvořena obrazem s vyšším měřítkem. Pohledem frekvenční analýzy jsou detaily v obraze reprezentovány nejvyššími frekvencemi. Spolu se zvyšováním měřítka tedy dochází k filtraci typu dolní propust a v obraze jsou tak zachovány jen frekvence nižší. Scale-space členíme na tzv. oktávy skládající se z vrstev o stejném rozměru, přičemž nejvyšší vrstva má dvojnásobné měřítko oproti první vrstvě. Základem následující oktávy je poslední vrstva oktávy předešlé, jež je podvzorkována na poloviční rozměr (viz Obr. 2). Scale-space je také označován jako měřítkově nezávislá forma obrazu.



Obrázek 2: Ukázka diskrétního scale-space.

Během tohoto kroku metody SIFT jsou nalezeny potenciaálně významné body (kandidáti), jenž zůstávají stabilní i během změny měřítka obrazu. Je tedy sestavena funkce, která postupně převádí vstupní obraz do různých měřítek a v takto vzniklých obrazech následně dochází k lokalizaci významných bodů. Jak bylo dříve dokázáno, jedinou možnou

funkcí splňující tento účel je funkce Gaussova. Je tedy zkonstruována tzv. Gaussova pyramida skládající se z obrazů $L(x, y, \sigma)$, které vznikají konvolucí Gaussova jádra $G(x, y, \sigma)$ o různém měřítku σ (šířka jádra) se vstupním obrazem $I(x, y)$:

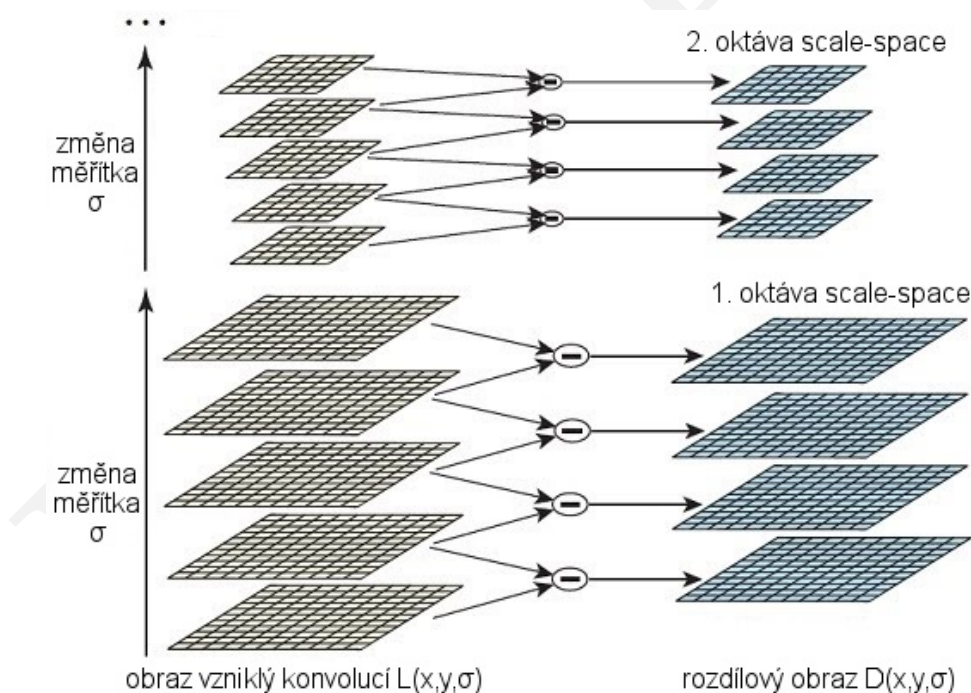
$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad \text{kde } G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x^2 + y^2)}{2\sigma^2}\right). \quad (12)$$

Při sestavování Gaussovy pyramidy se využívá vlastnosti soběpodobnosti Gaussovy funkce. To znamená, že vyšší vrstva je získána opakovanou aplikací konvoluce na vrstvu předešlou bez nutnosti neustálého zvětšování šířky konvolučního jádra.

Samotná scale-space, v níž jsou následně hledány extrémny, je sestavena z rozdílových obrazů $D(x, y, \sigma)$, které vznikají rozdílem dvou sousedících obrazů $L(x, y, \sigma)$ v Gaussově pyramidě. Je zřejmé, že sousedící obrazy mají rozdílné měřítko, které vyjadřuje koeficient k :

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma). \quad (13)$$

Obraz $D(x, y, \sigma)$ se nazývá rozdíl Gaussových funkcí. Konstrukci scale-space pomocí rozdílových obrazů znázorňuje Obr. 3.

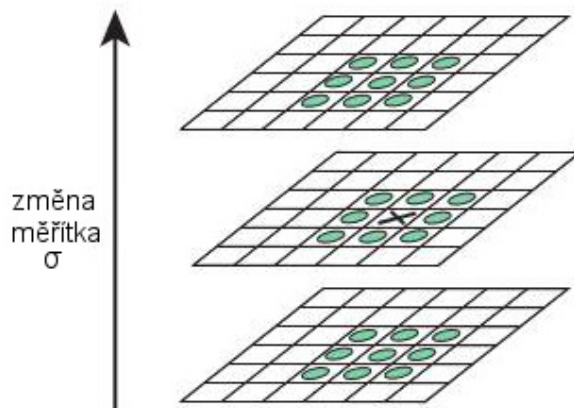


Obrázek 3: Konstrukce měřítkově nezávislé formy (scale-space) v případě metody SIFT.

Lokální maxima a minima v $D(x, y, \sigma)$ jsou nalezeny prozkoumáním 26-okolí každého bodu obrazu (viz Obr. 4). Pokud takový bod má nejmenší či největší hodnotu ve svém okolí, stává se kandidátem na významný bod objektu.

ZPŘESNĚNÍ POLOHY VÝZNAMNÝCH BODŮ

Kandidáti nalezení v předchozím kroku metody jsou charakterizovány svou souřadnicí a měřítkem. Pro tyto body je vhodné přesněji určit jejich polohu a naopak z množiny



Obrázek 4: Detekce lokálního extrému; pixel (označený X) je porovnáván s pixely ve svém 26-okolí (8+18 ze sousedících obrazů).

nalezených kandidátů vypustit takové body, které mají nízký kontrast (jsou citlivé na šum), nebo se nacházejí v blízkosti hran v obraze. Body ležící v blízkosti hran jsou špatně lokalizovatelné a jako takové, musejí být dodatečně odfiltrovány.

Poloha kandidáta (extrému) je aproximována kvadratickou funkcí. Za tímto účelem byla vyvinuta metoda, která pracuje na základě Taylorova rozvoje obrazu $D(x, y, \sigma)$, jehož počátek je posunut do zkoumaného bodu \mathbf{x} :

$$D(\mathbf{x}) = D + \frac{\partial D^\top}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^\top \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x}, \quad (14)$$

kde D je hodnota derivace ve zkoumaném bodě $\mathbf{x} = (x, y, \sigma)^\top$. Přesná pozice kandidáta je určena posunutím $\hat{\mathbf{x}}$, které odpovídá hodnotě extrému Taylorova rozvoje. Ten je spočten pomocí derivace funkce $D(\mathbf{x})$ podle \mathbf{x} a následným položením rovno nule

$$\hat{\mathbf{x}} = \frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2} \frac{\partial D}{\partial \mathbf{x}}. \quad (15)$$

Posunutí $\hat{\mathbf{x}}$ dosahující větších hodnot než 0,5 v kterékoliv dimenzi, značí skutečnost, že se extrém nachází blíže k sousednímu bodu. V takovém případě je zkoumaný bod nahrazen bodem sousedním a celý výpočet se opakuje. Pokud $\hat{\mathbf{x}}$ nepřesahuje hodnotu 0,5 je výsledné posunutí přičteno k souřadnicím zkoumaného bodu, čímž je získána sub-pixelová / sub-měřítková poloha extrému.

Odstranění bodů s nízkým kontrastem

Hodnota Taylorova rozvoje v bodě $\hat{\mathbf{x}}$ slouží k odhalení kandidátů s nízkým kontrastem a je získána pouhým dosazením $\hat{\mathbf{x}}$ z 15 do rovnice 14

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2} \frac{\partial D^\top}{\partial \mathbf{x}} \hat{\mathbf{x}}. \quad (16)$$

Takové kandidáty je totiž nutno vzhledem k jejich nestabilitě z množiny významných bodů odebrat. V normalizovaném obraze (jas v intervalu $\langle 0, 1 \rangle$) jsou za body s nízkým kontrastem považovány pixely mající hodnotu $|D(\hat{\mathbf{x}})|$ menší než 0,03 (hodnota empiricky zvolena).

Odstranění bodů v blízkosti hran

Ačkoliv rozdíl Gaussových funkcí $D(x, y, \sigma)$ vykazuje velkou odezvu v blízkosti hran obrazu, není vhodné takové pixely řadit do množiny významných bodů. Tyto body jsou totiž špatně lokalizovatelné a tudíž nestabilní.

Špatně definované extrémy funkce $D(x, y, \sigma)$ vykazují velkou hlavní křivost podél hran, avšak malou křivost v ortogonálním směru. Hlavní křivost v okolí zkoumaného bodu lze vyjádřit pomocí vlastních čísel Hessianovy matice:

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}. \quad (17)$$

Spočtení vlastních čísel je ovšem výpočetně náročná operace a proto se místo konkrétní hodnoty těchto čísel hledá jejich poměr. Pokud se totiž zkoumaný bod nachází v blízkosti hrany, je jeho jedno vlastní číslo, označené α , větší než to druhé, označené β , a současně platí, že

$$\text{Tr}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta, \quad (18)$$

$$\text{Det}(\mathbf{H}) = D_{xx}D_{yy} - D_{xy}^2 = \alpha\beta, \quad (19)$$

kde $\text{Tr}(\mathbf{H})$ je stopa Hessianu a $\text{Det}(\mathbf{H})$ jeho determinant. Nyní zavedeme r jako poměr mezi vlastními čísly, musí tedy platit rovnice $\alpha = r\beta$. Jednoduchou úpravou výše uvedených rovnic a dosazením lze získat tvar:

$$\frac{\text{Tr}(\mathbf{H})^2}{\text{Det}(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r}. \quad (20)$$

Lze tedy vypočítat, že poměr mezi stopou Hessianu $\text{Tr}(\mathbf{H})$ a jeho determinantem $\text{Det}(\mathbf{H})$ je závislý pouze na poměru vlastních čísel a ne na jejich přesných hodnotách. Proto postačí zavést jednoduchou podmínku porovnávající křivost zkoumaných bodů a určit prahovou hodnotu poměru r :

$$\frac{\text{Tr}(\mathbf{H})^2}{\text{Det}(\mathbf{H})} < \frac{(r + 1)^2}{r}. \quad (21)$$

Extrém funkce $D(x, y, \sigma)$ je zachován v množině významných bodů jen v případě, že je výše uvedená nerovnost splněna. Práh r je vstupním parametrem metody SIFT.

PŘIŘAZENÍ ORIENTACE VÝZNAMNÝM BODŮM

Body, které vzešly z předchozích dvou kroků metody SIFT, jsou nyní považovány za body významné. Jsou to body jednoznačně lokalizované ve scale-space a tedy invariantní vůči měřítku. Aby mohli být významné body považovány za invariantní i vůči rotaci, musí jim být přiřazena jejich dominantní orientace.

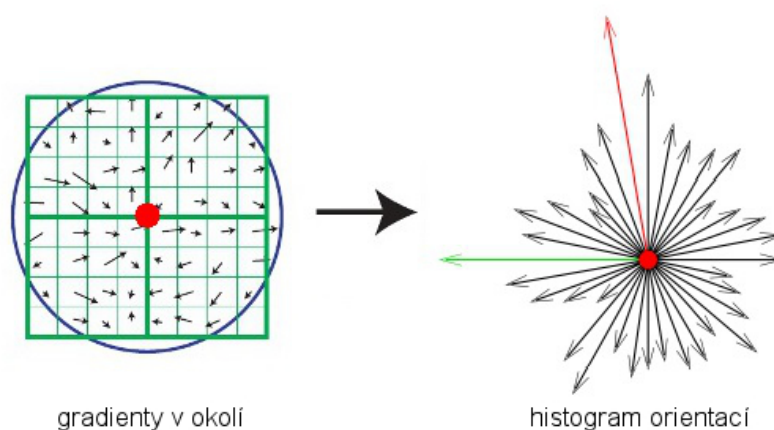
Proces přiřazení orientace probíhá tak, že nejdříve je ke zkoumanému bodu $\mathbf{x} = (x, y, \sigma)^\top$, $\mathbf{x} \in D(x, y, \sigma)$, nalezen měřítkově nejbližší obraz $L(x, y, \sigma)$ nacházející se v

již dříve sestrojené Gaussově pyramidě. Tímto přístupem je zajištěna měřítková nezávislost. Pro každý bod obrazu $L(x, y, \sigma)$, daného měřítka σ , je pak spočtena pomocí rozdílů jasových hodnot pixelů velikost gradientu $m(x, y)$ a orientace gradientu $\theta(x, y)$:

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2}, \quad (22)$$

$$\theta(x, y) = \tan^{-1} \left[\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right]. \quad (23)$$

Následně je zkonstruován histogram orientací sestávající se z gradientů spočtených v okolí významného bodu. Histogram orientací má 36 binů pokrývajících rozsah rotace o 360° . K binům histogramu s odpovídající orientací jsou přičteny hodnoty vážené velikostmi gradientů a koeficienty Gaussova kruhového okna (viz Obr. 5). Gaussovo okno má střed v zkoumaném bodu a její σ odpovídá 1,5 násobku měřítka významného bodu.



Obrázek 5: Gradienty spočtené v okolí významného bodu (vlevo) vážené Gaussovým oknem (modrý kruh) a sestavený histogram (vpravo) s více dominantními orientacemi (červená a zelená šipka).

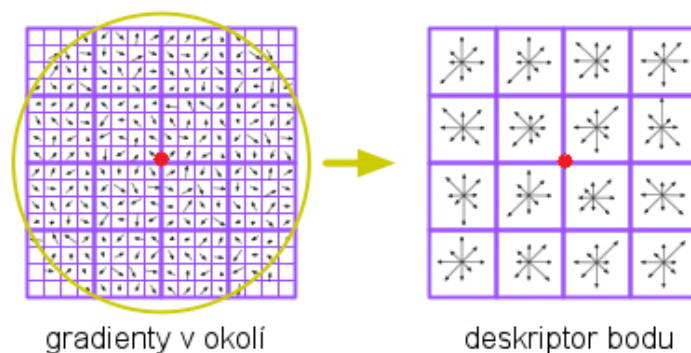
Dominantní orientace významného bodu je poté určena globálním maximem v sestaveném histogramu orientací. Pokud se ovšem v histogramu vyskytují lokální maxima, které dosahují alespoň 80% hodnoty globálního, označuje se takový histogram za histogram s více dominantními orientacemi. V takovém případě je na stejných souřadnicích vytvořen další významný bod s orientací odpovídající druhé (třetí, čtvrté, ...) nejvíce zastoupené orientaci v histogramu.

SESTAVENÍ DESKRIPTORU VÝZNAMNÝCH BODŮ

V předchozích krocích byla nalezena množina kandidátů na významné body v obraze. Z této množiny byly postupně odstraněny body vykazující nestabilitu a u zbývajících byla přesně definována jejich poloha. Dále byla jednotlivým významným bodům přiřazena dominantní orientace pomocí konstrukce histogramu orientací v jejich okolí. Nyní do celého procesu vstupuje tzv. deskriptor. Jeho úkolem je popsat okolí jednotlivých významných

bodů, tak aby tento popis byl nezávislý na geometrických a jasových transformacích obrazu.

Deskriptor se sestavuje z gradientů v blízkém okolí významného bodu. Je tedy množné pro jeho konstrukci využít výsledky z předešlých kroků metody. Nejprve je významnému bodu vybrán nejbližší obraz $L(x,y)$ z Gaussovy pyramidy, tak aby byla opět zachována nezávislost na měřítku. Okolí zkoumaného bodu je rozděleno na 8 stejně velkých čtvercových oblastí a pro každou oblast je sestaven 8-binový histogram orientací. Tyto histogramy jsou poté natočeny podle určené dominantní orientace významného bodu, čímž je zajištěna nezávislost na rotaci. Rozčlenění okolí do několika oblastí má výhodu v tom, že výsledný deskriptor je odolný vůči malým posunům obrazu. Za účelem zesílení této odolnosti SIFT navíc používá trilineární interpolaci, která výslednou hodnotu gradientu rozkládá mezi sousední biny v histogramu orientací. Výsledný deskriptor se skládá z jednotlivých histogramů nacházejících se v okolí významného bodu (viz Obr. 6) a je označován jako 128-binový deskriptor (okolí obsahuje 4x4 oblastí a každá obsahuje 8-binový histogram).



Obrázek 6: Výsledný deskriptor metody SIFT (vpravo) okolí významného bodu.

Nezávislost deskriptoru na transformacích jasové funkce obrazu lze jednoduše dokázat. Jasové transformace obrazu mohou být vyjádřeny následujícím vztahem

$$G(p) = AF(p) + B, \quad (24)$$

kde $G(p)$ a $F(p)$ je nová, respektive původní jasová funkce obrazu, parametr A rozhoduje o kontrastu a B pak o hodnotě jasu. Pokud tedy dojde ke změně osvětlení (zvýšení/snížení jasu) přičte se k jednotlivým pixelům konstantní hodnota, která se ovšem v gradientu nikterak neprojeví. Změna kontrastu se provede přenásobením hodnot jasu v obraze určitou konstantou, dojde tedy k nárůstu velikosti gradientu. Tento vliv lze jednoduše vyloučit normalizací vektoru deskriptoru. Získaný deskriptor metody SIFT je po normalizaci odolný i vůči jasovým transformacím obrazu.

3 SURF

Metoda SURF (z ang. *Speeded-Up Robust Features*) byla představena roku 2006 a jejím hlavním záměrem byla konstrukce výpočetně rychlého a stabilního deskriptoru schopného pracovat v reálném čase. Autoři se v mnohém inspirovali principy SIFT, které vykazovaly velice dobré výsledky. Zrychlení oproti SIFT autoři dosáhli v oblasti sestavení měřítkově nezávislé reprezentace obrazu (scale-space) a v redukci velikosti výsledných deskriptorů.

DETEKCE VÝZNAMNÝCH BODŮ

SIFT scale-space aproximuje pomocí rozdílů Gaussových funkcí. Ty ovšem mají velké odezvy i v okolí hran obrazu, které je potřeba z množiny významných bodů dodatečně odstranit pomocí Hessianu. Metoda SURF tyto dva kroky slučuje a významné body detekuje přímo pomocí determinantu Hessianovy matice. Rychlý výpočet determinantu je umožněn aplikací integrálního obrazu.

Integrální obraz

Integrální obraz je struktura vybudovaná nad vstupním obrazem umožňující rychlé spočtení součtu hodnot uvnitř libovolné obdélníkové oblasti vstupního obrazu. Může být vyjádřen následujícím předpisem

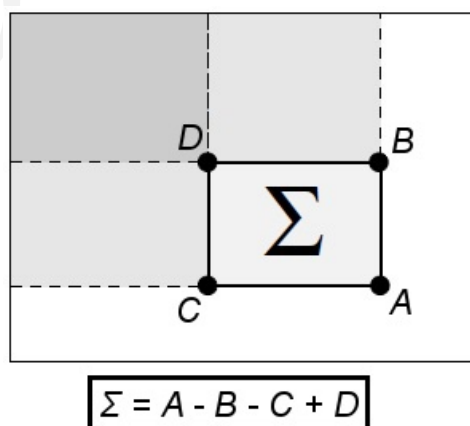
$$I_{\Sigma}(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j), \quad (25)$$

kde $I(i, j)$ je vstupní obraz a $I_{\Sigma}(x, y)$ představuje obraz integrální.

Součet hodnot obdélníkového regionu uvnitř vstupního obrazu lze pak jednoduše vyčíslit dosazením do vztahu:

$$\Sigma = A - B - C + D, \quad (26)$$

kde Σ značí hledaný součet a A, B, C, D jsou hodnoty integrálního obrazu v daných souřadnicích (viz Obr. 7).



Obrázek 7: Výpočet součtu hodnot v regionu Σ pomocí integrálního obrazu.

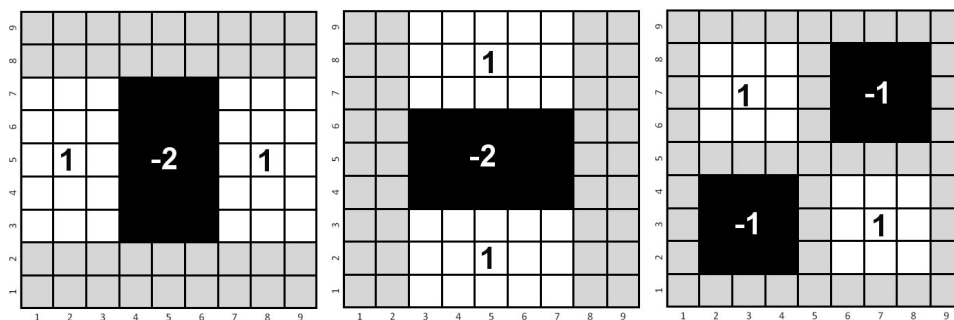
Hessianova matice

Hessianova matice bodu (x, y) v daném měřítku σ je definována jako

$$H(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{yx}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix}, \quad (27)$$

kde $L_{xx}(x, y, \sigma)$ je hodnota druhé parciální derivace podle x konvoluce vstupního obrazu $I(x, y)$ s Gaussovou funkcí $G(x, y, \sigma)$ s měřítkem σ v bodě (x, y) .

Metoda SURF pro výpočet prvků Hessianovy matice využívá obdélníkové funkce. Tyto funkce aproximují druhou derivaci v diskrétní formě a jedná se tak o konvoluci obrazu s filtry znázorněnými na Obr. 8. Je zřejmé, že implementace integrálního obrazu umožňuje velice rychlé spočtení odezvy na dané obdélníkové filtry.



Obrázek 8: Aproximace druhé derivace Gaussovy funkce podle x , podle y a podle xy (šedá pole mají 0 hodnotu).

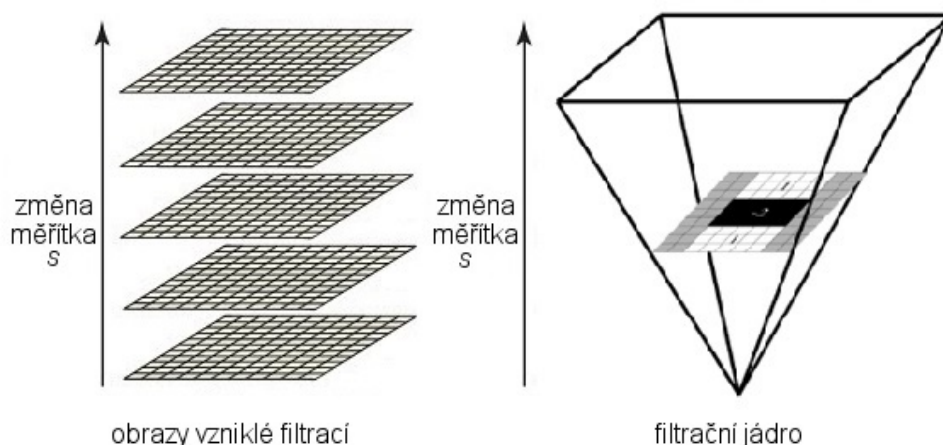
Znázorněné filtry mají velikost 9×9 pixelů a odpovídají nejmenšímu uvažovanému měřítku během sestavování scale-space metody SURF. Obrazy vzniklé filtrací jsou označovány jako D_{xx} , D_{yy} a D_{xy} a determinant Hessianu je spočten podle vzorce

$$\text{Det}(\mathbf{H}_{\text{aprox}}) = D_{xx}D_{yy} - (wD_{xy})^2. \quad (28)$$

Odezvy na obdélníkové filtry jsou dále normalizovány vzhledem ke své velikosti, což umožňuje ponechat váhu w konstantní pro všechna měřítka. Tato váha je použita pro co nej-
přesnější aproximaci Gaussova jádra a byla empiricky stanovena na hodnotu 0,9.

Sestavení scale-space

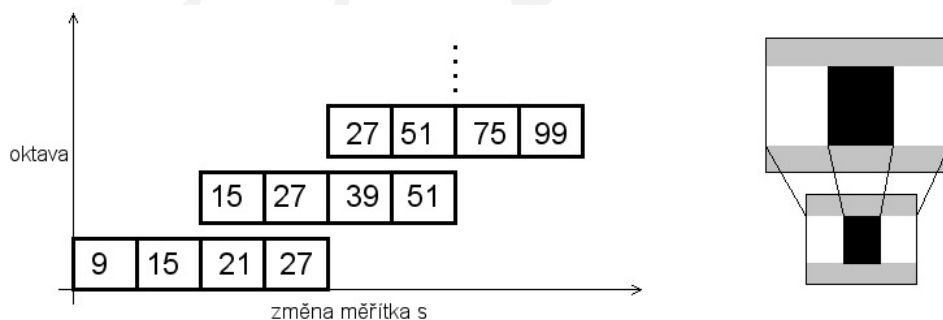
Významné body musejí vykazovat stabilní detekci i přes změnu měřítka obrazu, proto je nutné takové body lokalizovat uvnitř měřítkově nezávislé reprezentace. V případě metody SURF je scale-space zkonstruován z jednotlivých obrazů vzniklých filtrací. Díky implementaci integrálního obrazu nedochází k nárůstu výpočetní složitosti při zvětšení filtračního jádra. Tato složitost je totiž konstantní. Z tohoto důvodu se větších měřítek ve scale-space dosahuje konvolucí vstupního obrazu s filtračním jádrem zvyšujících se rozměrů. Odpadá tedy nutnost kaskádovitě generování scale-space, čímž je dosaženo ještě většího zrychlení oproti metodě SIFT. Obrazy navíc není potřeba podvzorkovat, což znamená, že nedochází k tzv. aliasingu.



Obrázek 9: Princip konstrukce scale-space metody SURF.

Scale-space se dělí na oktávy. Základem první oktávy (její první vrstvou) je obraz vzniklý aplikací filtračních jader o rozměru 9×9 . Pro zvyšující se měřítko filtrační jádra postupně narůstají na rozměr 15×15 , 21×21 a 27×27 . Pro tento rozměr jádra je již splněna podmínka, že nejvyšší vrstva oktávy má dvojnásobné měřítko nežli vrstva prvotní. Krok o který jednotlivá jádra narůstají, je označován jako l_0 a byl stanoven na $2/3$ velikosti základního jádra. Platí tedy, že $l_0 = 6$.

Základem následující oktávy je vždy druhá vrstva oktávy předešlé a zároveň je při jejím sestavování krok l_0 , použitý při konstrukci předešlé oktávy, zdvojnásoben. Druhou oktávu tedy tvoří odezvy na filtrační jádra o velikosti 15×15 , 27×27 , 39×39 , 51×51 ($l_0 = 2 \cdot 6 = 12$).

Obrázek 10: Změny rozměrů filtračních jader pro jednotlivé oktávy scale-space (vlevo) a názorná ukázka změny rozměru jádra (vpravo). Poznamenejme, že krok l_0 je vždy sudý (6, 12, 24) tak, aby při zvyšování měřítka nedocházelo ke změně struktury filtračních jader.

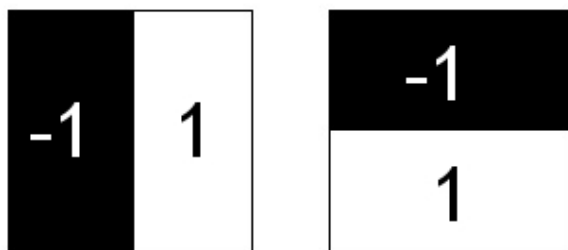
Jak již bylo zmíněno výše, SURF využívá k detekci významných bodů determinant Hessiovny matice. Z výsledných odezev na filtraci je tedy postupně spočtena hodnota tohoto determinantu, čímž je získán obraz $H_{\text{Det}}(x, y, \sigma)$. Právě tyto obrazy představují finální měřítkově nezávislou reprezentaci vstupního obrazu, ve které jsou následně lokalizovány významné body.

Detekce maxim uvnitř scale-space

Významné body jsou detekovány jako maxima ve scale-space. Tyto maxima jsou nalezena stejným způsobem, který byl zvolen pro metodu SIFT. Body jsou tedy porovnávány se svým 26-okolím v rámci scale-space (8 sousedů + 18 sousedů z vyšší a nižší vrstvy). Pokud má daný bod ve svém okolí nejvyšší hodnotu, je považován za bod významný. Jeho lokalizace je následně zpřesněna pomocí Taylorova rozvoje (popsáno v kapitole SIFT), což vede k určení sub-pixelové a sub-měřítkové polohy významného bodu.

PŘIŘAZENÍ ORIENTACE VÝZNAMNÝM BODŮM

Metoda SURF se, podobně jako metoda SIFT, snaží přidělit každému významnému bodu $\mathbf{x} = (x, y, \sigma)^\top$ jeho dominantní orientaci, tak aby výsledný deskriptor byl invariantní vůči rotaci. V případě SURF ovšem není využíván histogram orientací, nýbrž se zkoumá odezva na tzv. Haarovu vlnku v kruhovém okolí daného bodu. Toto kruhové okolí má střed ve zkoumaném bodě a poloměr $6s$, kde s značí měřítko vrstvy ve scale-space, v kterém se významný bod nachází. Haarova vlnka je opět aproximovaná pomocí obdélníkových filtrů ve směru osy x a y (viz Obr. 11), což umožňuje zapojení integrálního obrazu do procesu výpočtu odezvy. Tyto filtry mají opět měřítkově závislou velikost $4s \times 4s$ a krok jejich posunu v okolí bodu byl stanoven na s .



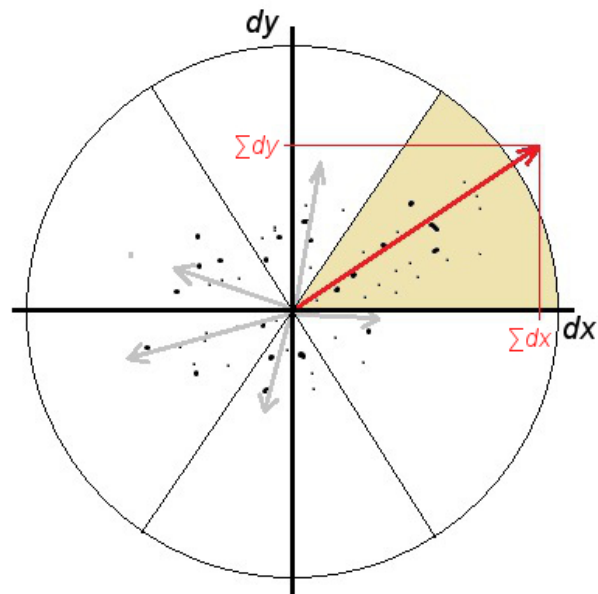
Obrázek 11: Haarova vlnka aproximovaná obdélníkovými filtry ve směru osy x a y .

Výsledky konvoluce obrazu s Haarovou vlnkou se označují jako d_x a d_y . Tyto odezvy jsou, stejně jako u SIFT, váženy koeficienty Gaussova kruhového okna se středem ve zkoumaném bodě a parametrem $\sigma = 2,5s$. Vážené d_x a d_y jsou následně zakresleny do souřadnicového systému (viz Obr. 12) a výsledná dominantní orientace významného bodu je určena jejich součtem v rámci výseče pokrývající $\pi/3$ prostoru. V těchto výsečích tedy vznikají vektory definované Σd_x a Σd_y a výseč, ve které se nachází ten největší z nich, představuje dominantní orientaci významného bodu.

SESTAVENÍ DESKRIPTORU VÝZNAMNÝCH BODŮ

Prvním krokem ke konstrukci deskriptoru je vytvoření čtvercové oblasti obklopující významný bod. Tato oblast se středem ve zkoumaném bodě je natočena podle dominantní orientace daného bodu a její hrana má délku $20s$ (s je měřítko ve scale-space).

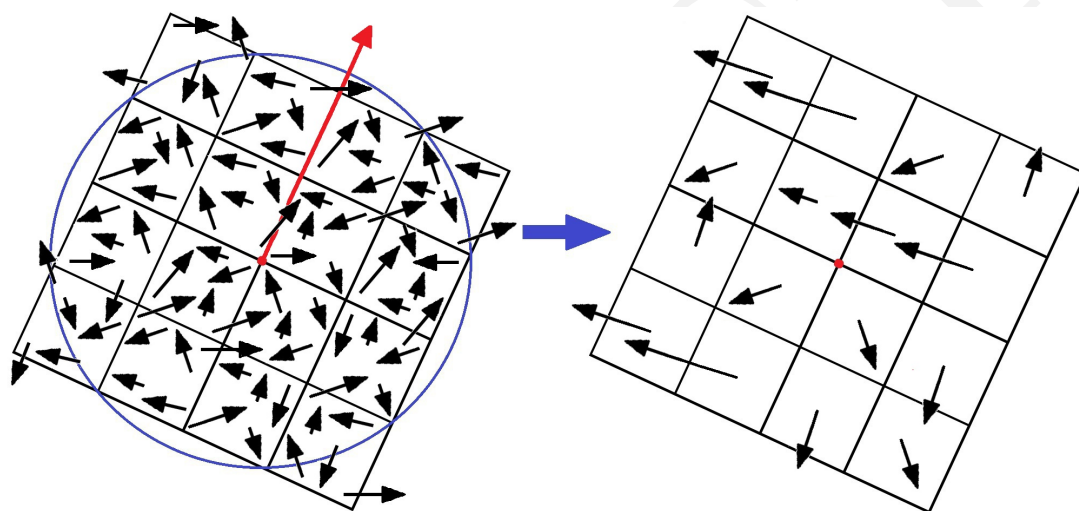
Čtvercová oblast je dále rovnoměrně rozdělena na 4×4 podoblasti. V každé podoblasti je určeno pět pravidelně rozmístěných bodů, pro které je vypočtena jejich odezva na



Obrázek 12: Určení dominantní orientace metodou SURF; největší vektor (vyznačen červeně) se nachází ve výseči (žlutá barva), která určuje výslednou orientaci významného bodu.

Haarovou vlnku ve směru osy x a y (souřadnicový systém je určený dominantní orientací významného bodu). Odezvy této pětice bodů jsou získány aplikací obdélníkových filtrů (viz Obr. 11), jako tomu bylo v předchozím kroku metody SURF, s tím rozdílem, že nyní je rozměr filtračních jader stanoven na $2s \times 2s$ a není možné využít integrální obraz. Odezvy jsou značeny d_x a d_y a pro zvýšení robustnosti výsledného deskriptoru jsou opět váženy koeficienty Gaussova kruhového okna se středem ve významném bodě a parametrem $\sigma = 3, 3s$.

V rámci každé podoblasti je poté spočtena Σd_x a Σd_y . Z důvodu zvýšení odolnosti deskriptoru na změny osvětlení v obraze, jsou zjištěny i hodnoty $\Sigma |d_x|$ a $\Sigma |d_y|$. Tyto čtyři hodnoty pak tvoří vektor $v = (\Sigma d_x, \Sigma d_y, \Sigma |d_x|, \Sigma |d_y|)$ pro každou podoblast v okolí významného bodu. Výsledný deskriptor (viz Obr. 13) se skládá z takto získaných vektorů pro všech 16 podoblastí a okolí významného bodu tak popisuje pomocí 64 binů (16 podoblastí \times 4-prvkový vektor v).



Obrázek 13: Odezvy na Haarovu vlnku v okolí významného bodu (vlevo) podle dominantní orientace (červená šipka) a výsledný deskriptor metody SURF (vpravo).

4 Zkouška

4.1 Harris Corner Detection

4.1.1 Co to je?

Metoda hledání rohů v obraze.

4.1.2 Na co to je?

Metda najde v obraze rohy, které jsou významné body. To nám umožní kompaktně reprezentovat objekty v obraze, pro popis, rozpoznávání, sledování (tracking)

4.2 SIFT

4.2.1 Co to je?

Scale Invariant Feature Transform - metoda pro detekci klíčových bodů.

4.2.2 Na co to je?

Metoda slouží k detekci významných bodů v scale prostoru. Ke každému bodu je určené měřítko a orientace. Pak lze vytvořit deskriptor, který je nezávislý na měřítku a orientaci.

4.3 SURF

4.3.1 Co to je?

Speed-Up Robust Features - metoda pro detekci klíčových bodů.

4.3.2 Na co to je?

Stejně jako SIFT - ale využívá aproximace a integrální obraz pro urychlení.

Reference

DRAFT